AFRL-RI-RS-TR-2009-94
Final Technical Report
April 2009

HIGH PERFORMANCE COMPUTING INNOVATION SERVICE PORTAL STUDY (HPC-ISP)

University of Southern California

Sponsored by Defense Advanced Research Projects Agency
DARPA Order No. AD94/00

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

STINFO COPY

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications, or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the 88th ABW, Wright-Patterson AFB Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (http://www.dtic.mil).

AFRL-RI-RS-TR-2009-94 HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION IN ACCORDANCE WITH ASSIGNED DISTRIBUTION STATEMENT.

FOR THE DIRECTOR:

/s/ /s/

CHRISTOPHER J. FLYNN
Work Unit Manager

EDWARD J. JONES, Deputy Chief
Advanced Computing Division
Information Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
APR 09	Final	Aug 06 – Sep 08

4. TITLE AND SUBTITLE

HIGH PERFORMANCE COMPUTING INNOVATION SERVICE PORTAL STUDY (HPC-ISP)

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA8750-06-1-0240

5c. PROGRAM ELEMENT NUMBER
62303E

6. AUTHOR(S)

Brian Schott
Suzy Tichenor
Robert Graybill

5d. PROJECT NUMBER
AD94

5e. TASK NUMBER
IS

5f. WORK UNIT NUMBER
IE

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Southern California
Information Sciences Institute
4676 Admiralty Way
Marina Del Rey CA 90292-6695

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

AFRL/RITB
525 Brooks Rd.
Rome NY 13441-4505

Defense Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington VA 22203-1714

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-RI-RS-TR-2009-94

12. DISTRIBUTION AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED. PA# 88ABW-2009-1288

13. SUPPLEMENTARY NOTES

14. ABSTRACT

This project was a study by the Information Sciences Institute (ISI), the Council on Competitiveness (Council), Pratt & Whitney (P&W), Ohio Supercomputer Center (OSC), and Georgetown University (GU) intended to: 1) identify why companies that do not currently employ HPC for advanced modeling and analysis have failed to adopt this technology when the benefits have been showcased so compellingly, and 2) develop technical and business concepts that could help enable these "desktop-only" users to employ more advanced computing solutions in their manufacturing design cycles. The products of this study include: A broad industry survey of desktop and entry-level HPC users "Council on Competitiveness and USC-ISI Study of Desktop Technical Computing End Users and HPC", an in-depth industry user survey "Council on Competitiveness and USC-ISI In-Depth Study of Desktop Technical Computing End Users", a case study of Advanced Computational and Engineering Services (ACES), "ACES: Not Ready To Play the HPC Card", a case study of Woodward, a supplier of P&W, "Woodward FST: Software Costs and Finding Experts Are Stalling HPC Adoption", and a business and technical concept study "Innovation – Solution Portal to Information, Resources, and Experts (I SPIRE^{SM})".

15. SUBJECT TERMS
High Performance Computing, Multi-Core Computer Architecture, Parallel Computing

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Christopher J. Flynn
U	U	U	UU	27	19b. TELEPHONE NUMBER (*Include area code*) N/A

Table of Contents

List of Figures

1. Summary

The following is the final report for the project entitled High Performance Computing Innovation Service Portal Study (HPC-ISP). This project includes independent studies related to high performance computing challenges. The following final report covers the full period of performance of this effort, from 8/21/2006 to 9/21/2008.

1.1 High Performance Computing – Innovation Service Portal (HPC-ISP) Phase 1 Study

The *High Performance Computing – Innovation Service Portal (HPC-ISP) Phase 1* effort was a nine-month study by the Information Sciences Institute (ISI), the Council on Competitiveness (Council), Pratt & Whitney (P&W), Ohio Supercomputer Center (OSC), and Georgetown University (GU) whose intent was to: 1) identify why companies that do not currently employ HPC for advanced modeling and analysis have failed to adopt this technology when the benefits have been showcased so compellingly, and 2) develop technical and business concepts that could help enable these "desktop-only" users to employ more advanced computing solutions in their manufacturing design cycles.

The products of *Phase 1* delivered to DARPA include:

A broad industry user survey "Council on Competitiveness and USC-ISI Study of Desktop Technical Computing End Users and HPC" [13],

An in-depth industry user survey "Council on Competitiveness and USC-ISI In-Depth Study of Desktop Technical Computing End Users" [14],

A case study of Advanced Computational and Engineering Services (ACES), "ACES: Not Ready To Play the HPC Card",

A case study of Woodward, a supplier of P&W, "Woodward FST: Software Costs and Finding Experts Are Stalling HPC Adoption", and

A business and technical concept study "Innovation – Solution Portal to Information, Resources, and Experts (I-SPIRE^{SM})".

The results of this project led to a follow on project funded by DARPA entitled High Performance Computing Innovation Service Portal Pilots (HPC-ISP-PILOTS).

1.2 Computer Architecture Industry-Academic Consortium (CAIAC) Study

In addition, a six month study lead by ISI entitled Computer Architecture Industry Academic Consortium (CAIAC) was performed to investigate long-range research into parallel systems, software, and applications for next generation computing devices. In this study, ISI organized three regional academic workshops that assembled leading professors from computing fields to investigate the multicore parallelism challenge. The consensus of the participants was that the computer industry is rapidly approaching a crisis where the future growth of processor performance for many applications is uncertain. The participants warned that a major research investment is required to "break the habit of von Neumann sequential computing models" in order to derive performance growth from chip-level parallelism. They recommended that the government and industry fund "diverse academic research efforts" that are able to break traditional hardware and software boundaries in the development of new parallel software systems.

Using this guidance, ISI brought together DARPA representatives with leading industry executives from computing vendors, software vendors, and computing users to explore the potential of industry-government consortium approaches for funded focused research in this area. A concept paper, entitled *Parallel Systems, Software, and Applications Research (PSSAR) Consortium Concept Paper*, was submitted to DARPA.

2. High Performance Computing – Innovation Service Portal (HPC-ISP) Study

2.1 Introduction

Virtual prototyping and large-scale data modeling through technical computing plays a key role in the design and manufacture of many industrial products ranging from jet engines to potato chips as well as industrial business processes such as energy exploration, financial modeling, and supply chain management. Given the complexity of these applications and the expanding range of high-value economic activities driven by technical computing, the market should naturally migrate to more productive high performance computing (HPC) solutions. However, this has not been the case. HPC represents only 3% of the overall computing market and is a small percentage of technical computing software sales. Despite the proven productivity and competitiveness gains that some companies have realized by using HPC, many U.S. industrial companies have not made the transition to this important technology. They appear to be stalled on the desktop. The "missing middle" of HPC users is shown in Figure 1.

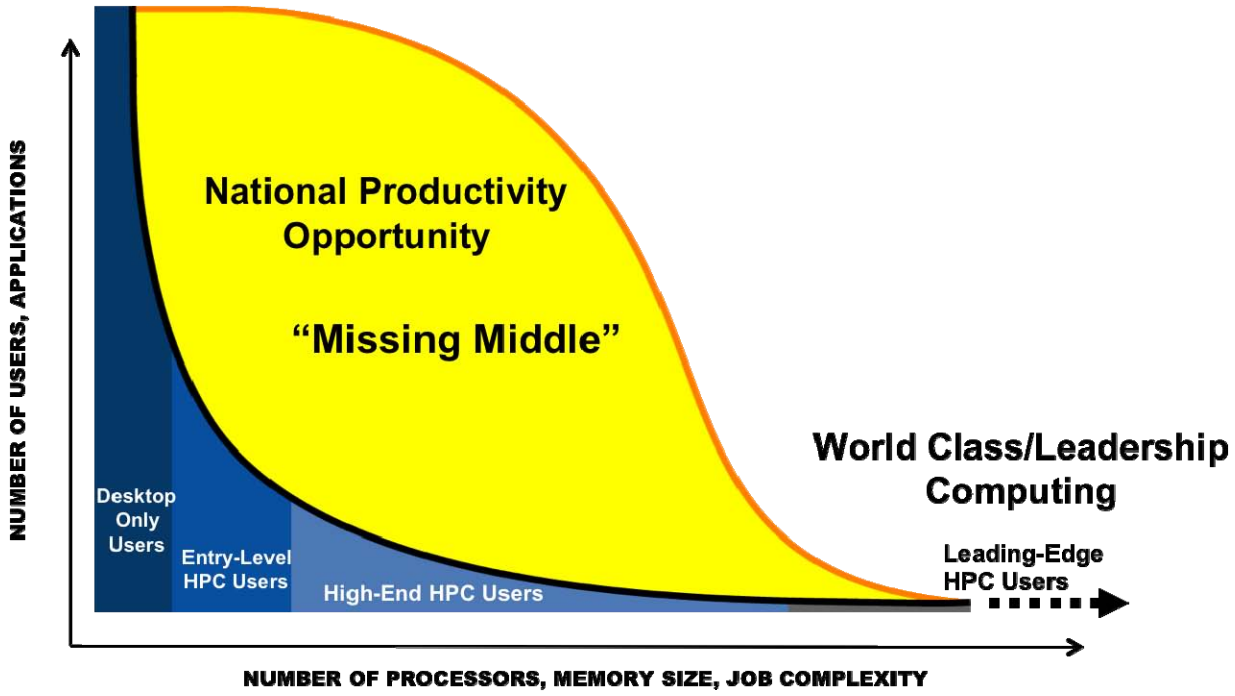


Figure 1. "Missing Middle" in the High Performance Computing Market

This "missing middle" in the HPC market is eroding U.S. economic competitiveness and is a threat to the superiority of DoD systems. The supplier base for major DoD systems, especially suppliers that provide subsystems and components, are threatened by global competition. It is essential that these suppliers remain competitive and maintain their technological advantage. In this increasingly competitive global environment, ***the country that wants to out-compete must be able to out-compute through making HPC-driven modeling and analysis a "best business practice."*** The goal of the study was to stimulate industrial innovation through expanded access to HPC.

2.2 Methods, Assumptions, and Procedures

The HPC-ISP approach was to define the business case and concepts for a national service that reaches out to the HPC "never-evers" – less experienced users that historically have been the economic engine of the U.S. and are now threatened by global competition. This study was organized into two distinct tasks as shown in Figure 2.

- Task 1 included a number of customer surveys and case studies to help understand the market forces and user dynamics that often determine success or failure irrespective of technical merit.
- Task 2 included a number of business and technical concept studies and identification of two pilot projects for HPC-ISP services. Together, these tasks resulted in a credible path forward that will enable a whole new industry community to increase their competitiveness through the use of an HPC collaboration ecosystem.

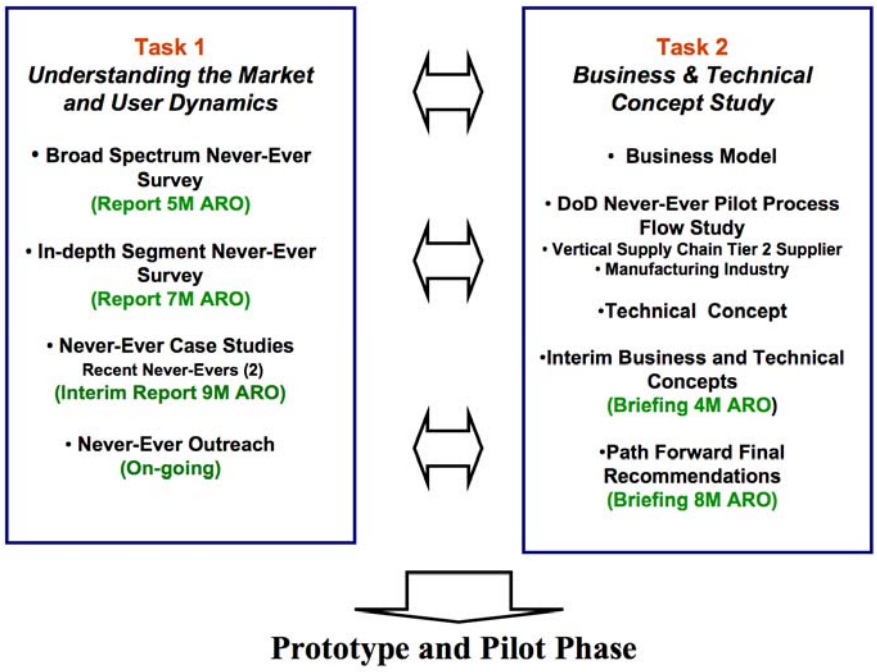


Figure 2. HPC-ISP Project Structure

2.2.1 Task 1: Understanding the Market and User Dynamics

Despite the proven productivity and competitiveness gains that some large companies have realized from using HPC, many firms within the U.S have "never ever" made the jump from technical desk top computing to this important technology. This user gap is particularly prevalent among small and medium sized companies, but does not exclude many well-known, large firms also. This activity examined: 1) why companies that do not currently employ HPC have failed to adopt this technology when the benefits have been showcased so compellingly, and 2) whether a National Innovation Collaboration Ecosystem with an Innovation Service Portal could help to rectify this situation. The goal was to better understand:

- Whether these "never ever" organizations are familiar with this technology and the benefits they could derive if they used it,
- What has prevented them from using it (Unfamiliarity? Cost of hardware/software? Talent required? Accessibility? Previous "bad" experience)
- What would motivate them to explore its capabilities,
- Willingness, ability and/or capacity to access such systems remotely at an off-site location such as a university, national laboratory or vendor operated center versus purchasing systems for in-house use,
- Each user's perception of the potential benefits these organizations could reap from its application, and
- How an HPC-ISP with a HPC collaboration backbone could meet their needs.
- How much are they willing to pay for these types of services and under what terms and conditions.

Some of the research mechanisms employed to gather this information included:

Broad Industry Survey of "Never Ever" Users. We developed and executed a national, cross industry survey of companies that are not using HPC in order to better understand the issues above. We engaged IDC, a market research firm knowledgeable about high performance computing technology and the current market, to conduct the survey. Council experience with industry surveys suggests that detailed telephone interviews provide a much higher response rate, as well as higher quality data than surveys conducted by mail or e-mail. A report with comprehensive quantitative and descriptive findings was published [[13]].

In-Depth Industry Segment Survey of "Never Ever" Users. We developed and executed a survey of companies from a specific industry segment that are not using HPC. The objective was to gain a more in-depth understanding of the issues above as they apply across an industry of interest to DARPA. A comprehensive report was published [14].

"Never-Ever" Case Studies. Using information from previous surveys, along with information gleaned from the Council's complementary HPC activities, we produced two case studies that focus on users that recently have migrated from technical desktop computing to small HPC clusters…in other words…from "never ever" to entry level HPC user. (HPC "never-evers" are defined as those users that currently are running technical applications only at the desktop. Less experienced/entry level HPC users are defined as having migrated from the desktop and currently running applications on small clusters up to around 64 nodes (typical ISV software application limit). These case studies explored the motivations that prompted these organizations to adopt HPC, the business and technical hurdles they had to address in the process, and the successes they achieved from using this technology. The intent was to capture the overall challenges, technical case, business case, lessons learned and recommendations for an improved ecosystem.

"Never-Ever Outreach. Based on the surveys performed above and related Council reports [1]-[11] a focused outreach effort and engagement with the broad HPC community was performed to test the concepts and business models developed under Task 2. This included media outreach, conference briefings (e.g. SuperComputing 2005, trade shows), and one-on-one customer visits.

2.2.2 Task 2: Business and Technical Concept Study

The fundamental findings and data derived from the customer-centric surveys throughout the nine-month study were used to develop candidate business models and proposed service-oriented architecture (SOA) for a high performance computing innovation service portal (web services portal), renamed to *Innovation – Solution Portal to Information, Resources, and Experts (I-SPIRE^{SM}).* The researchers selected representative DoD-relevant technology capability demonstration pilots and will map out the actual business, technical, and process or workflow requirements and issues that must be addressed if actual pilots were initiated in the next phase to test the business models. One technology demonstration pilot was anticipated to evaluate the opportunity to use a company's supply chain to broaden usage of HPC among "never-evers." In this scenario, an experienced HPC User, such as Pratt & Whitney, would identify a key supplier that does not currently use HPC, and work with them to use HPC to solve a specific problem. An independent software vendor and an HPC center will also be involved. A second pilot was expected to focus on a manufacturing segment or technology area. A potential manufacturing segment pilot example could include the Edison Welding Institute, Ohio Supercomputing Center, an ISV, and manufacturing company(s).

An interim business and technical model briefing was provided along with a recommended path forward briefing/review. The issues proposed by this study are of vital importance for the nation as it evaluates the costs and benefits of creating a national innovation collaboration ecosystem.

2.3 Results and Discussions

HPC-ISP completed two industry user surveys and two industry case studies. The studies included:

- The *Study of Desktop Technical Computing End Users and HPC* was a national cross-industry survey of 77 "desktop-only" companies varying in size from $1M to over $1B in revenue. The goal of the survey was to understand why a large group of companies that pursue technical computing on desktop systems have not advanced in greater numbers to HPC. Over half of the companies (57%) surveyed said that they have problems that they can't solve with existing desktop computers. A high proportion (53%) scale down their advanced problems, resulting in a loss of insight, innovation, and competitive gain. They identified ***three systemic barriers to HPC***: 1) ***lack of application software***, 2) ***access to talent***, and 3) ***cost constraints***. Over 29% would be willing to spend between $25K and $200K to explore the use of HPC.

- The *In-Depth Study of Desktop Technical Computing End Users* was a survey of 29 of the 250 Edison Welding Institute (EWI) members ranging in size from $10M to over $1B. The goal was to provide an in-depth understanding of HPC adoption issues for an industry relevant to the DoD (materials joining) and representative of other industries (concrete, polymers, etc.). The primary barriers to HPC adoption in this group were ***overall cost and return-on-investment justification***. The primary technical driver of HPC adoption for the desktop only companies within this group is the ***availability of "strategic fit" software*** (both models and applications) that address specific business requirements. Over 60% of firms would be willing to pay outside consulting organizations between $2K/month and $10K/month for realizing the benefits of HPC.

- In the case study of Advanced Computational and Engineering Services (ACES), *"ACES: Not Ready to Play the HPC Card"*, the goal was to understand the barriers to adoption that an experienced engineering

services company has encountered with HPC, and to determine what business and technical approaches would enable HPC use in the future. ACES is a five-person firm that provides a broad spectrum of design engineering services and consulting. It differentiates itself by tackling complex multidisciplinary problems by ***building custom-tailored software solutions*** that bridge single-discipline commercial software packages. ACES has HPC modeling and simulation expertise on staff, but is unable to utilize it due to the "stratospheric software costs." Other than payroll, software is ACES' largest business expense.

- In the case study of Woodward, *"Woodward FST: Software Costs and Finding Experts Are Stalling HPC Adoption"*, the goal was to understand the barriers that a desktop-only user in the DoD supply chain has encountered with HPC, and determine what would enable HPC use in the future. Woodward is an $80M, 230-person supplier of fuel injection systems for commercial and military system vendors such as Pratt & Whitney. Woodward designers scale down simulation models to fit on the desktop because of lack of HPC expertise, software, and computing. They are limited on the desktop to single-physics simulation, which leads to *turn-backs* and *design escapes.* Woodward lacks the IT expertise to maintain HPC systems or ***even to make a convincing return-on-investment (ROI) argument for HPC to management.***

2.3.1 Broad Industry Survey Results

"Council on Competitiveness and USC-ISI Study of Desktop Technical Computing End Users and HPC" was the first study ever aimed at understanding why the large percentage of companies that pursue technical computing on desktop systems (PCs, Macs, Unix workstations) have not advanced in greater numbers to using high performance computing (HPC). This national, cross-industry study surveyed 77 "desktop-only" companies that varied greatly in size and included many firms with more than 1,000 employees and over $1 billion in annual revenue. The surveyed companies were old hands at technical computing on the desktop, having practiced it for 16 years on average. They believe that technical computing is an important driver for their competitiveness and the vast majority (97%) of firms perform virtual prototyping or large-scale data modeling on their desktop computers in designing their products.

Over half of the study participants (57%) said that they have problems that they can't solve with their existing desktop computers. A high proportion (53%) were forced to scale down their advanced problems to fit their desktop computers, resulting in a loss of insight, innovation, and competitive gain. Others are choosing simply to ignore their advanced problems, with more potentially serious competitiveness consequences. A third strategy, pursued by more than half the firms, was to increase the amount of slower, more expensive physical prototyping. Previous Council on Competitiveness studies showed that these alternatives render companies more vulnerable to competitors that have greater determination to employ HPC servers for their proven benefits.

The study revealed that three systemic barriers are stalling HPC adoption among the desktop only users:

Lack of application software. The importance for industry of software applications used to model products and processes can hardly be overstated. The users of single-processor desktop computers in this study expressed strong concern about the availability of software that could run their problems on multiprocessor HPC servers.

Access to sufficient talent. For the desktop users, access to in-house or external HPC experts is another important prerequisite for HPC adoption. Lack of an adequate number of people skilled in using HPC hardware and software systems to run specific business and industrial problems was another frequently cited barrier.

Cost constraints. The third major barrier to HPC adoption was cost. Including the aforementioned talent and software costs was the difficulty of convincing the executive suite to approve HPC budgets and purchases.

These same issues have persisted as significant barriers to (expanded) HPC usage since the initial *"Council on Competitiveness Study of U.S. Industrial HPC Users"* (July 2004), and constitute serious deterrents to HPC adoption for the entire category of "desktop-only" technical computing companies. The implications for these firms, U.S. industry and the nation are sobering.

HPC can be a game-changing technology. Companies that fail to embrace it where appropriate may erode their competitive standing. The advanced problems their desktops cannot solve today represent lost opportunities for these firms to power ahead of their global competitors.

Large companies that use only desktop technical computing may be at even greater competitive risk than smaller firms that have not yet adopted HPC technology. Larger firms often are slower to adopt new and/or different technologies because their current technologies are deeply embedded into complex workflows that can be difficult and costly to revamp.

Critical supply chains and the leadership of many U.S. industries are also at risk if larger numbers of "desktop-only" firms do not advance to HPC-based modeling and simulation. If over time "desktop-only" firms cannot meet more complex requirements — and meet them at a faster pace — they will place themselves, as well as customers who rely on them, in competitive jeopardy.

Despite the potential competitive advantage associated with adopting HPC and the potential competitive disadvantage from failing to adopt the technology, the status quo uncovered by the study can be summed up as follows: About a third (34%) of the companies that have never used HPC do see a fit but are stalled by one or more major barriers. More than half (56%) believes their current approaches are adequate, and only a small minority (10%) anticipates adopting entry-level HPC systems (see Figure 3). Clearly an "enabling" function of some kind will be required to help "desktop-only" users transition to entry-level HPC systems. However, HPC does have perceived value in the technical computing market. Over 29% of the surveyed companies in the broad industry survey stated that they would be willing to spend between $25,000 and $200,000 to explore the use of HPC to improve their design and manufacturing processes.

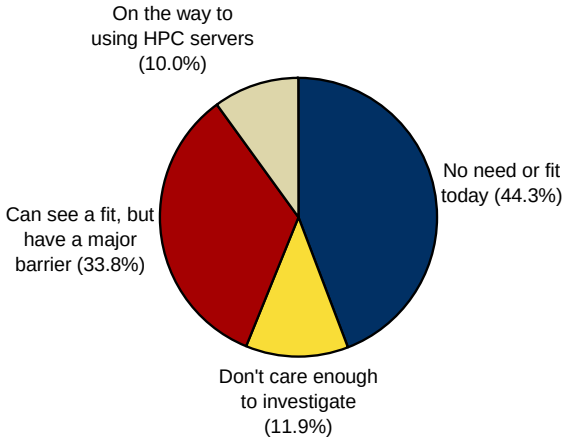


Figure 3. Broad Industry Survey Results: Participants' Views of HPC Server Fit

2.3.2 In-Depth Industry Survey Results

"Council on Competitiveness and USC-ISI In-Depth Study of Desktop Technical Computing End Users" provides a more in-depth understanding of the issues as they apply to an industry of interest to the DoD. Following the broad industry study, this in-depth study "drilled down" into a predefined group of desktop-only and entry-level HPC users to identify any significant differences from the broader study. The study participants consisted of members of the Edison Welding Institute (EWI), which is located in Columbus, Ohio. Twenty-nine of the approximately 250 member organizations participated, representing aerospace, automotive, energy, chemical, and heavy manufacturing industries. The surveyed companies varied greatly in size. About 62% claimed annual revenues exceeding $1 billion. About half used desktop systems exclusively for technical computing such as welding and joining process simulation, but 20% also used HPC servers, and another 24% were outsourcing HPC-based work to EWI or similar engineering services companies. Nearly all of the surveyed companies (96%) performed physical testing and prototyping and the percentage of companies doing at least some virtual prototyping was also high (81%). However, the percentage of physical prototyping performed (71%) is well over twice that of virtual prototyping (29%) in the survey.

There were several key similarities between the broad and in-depth studies:

There was no strong correlation between the size of a company (revenue, number of overall employees, number of technical employees) and whether a company was using HPC servers. As with the broad industry survey above, large firms are as likely as small companies to be stalled on desktop technical computers, at least where welding is concerned.

Although most respondents said their firms have large numbers of technical employees (scientists, engineers, analysts), there was no discernible correlation between the number of technical employees and whether the firms use HPC servers. It appears that while HPC has been used for some time in the R&D areas of many companies it is only beginning to be used in the manufacturing area for tasks such as welding. A company may have many technical employees who rely on HPC, but few if any of these employees may be associated with welding-related work.

Most of the companies in this study also had problems that they could not solve. As with the study participants in the broad industry survey, these firms are either ignoring those problems, scaling down the problems to fit their systems, or relying more heavily on slower and more expensive physical prototyping.

Participants in the in-depth study distinguished more clearly between business and technical drivers/motivators for adopting more powerful systems. The ***primary technical driver*** of HPC adoption for the desktop only companies within this study group is the availability of "strategic fit" software (both models and applications) that closely matches the specific problems the companies want to tackle on HPC systems. The ***principal business motivators*** are future (48%) and current (29%) customer requirements, followed by the need to catch up if competitors are using HPC (19%). This suggests that organizations with strong market leverage could guide their suppliers to adopting HPC through their requirements process.

If you were to use outside consulting organizations, which would you be most likely to use?

	Number of Responses			
	Rated #1	Rated #2	Rated #3	Total
A nonprofit organization like EWI	9	11	4	**24**
An engineering services company	4	5	5	**14**
A major university	7	3	2	**12**
A small system vendor that understands our needs	5	2	3	**10**
A government national laboratory including NSF centers	3	1	3	**7**
A smaller university	1	1	3	**5**
A large system vendor like IBM, HP, Sun, etc.	3	1	0	**4**
Local technical experts	1	1	2	**4**
ISV application software provider	0	1	1	**2**
A community college or technical school	0	0	1	**1**

Figure 4. In-Depth Survey Results: Use of Outside Consulting Organizations

Study participants also made clearer distinctions between "drivers/motivators" and the barriers preventing HPC adoption. The availability of strategic fit software was seen as "motivator" and cost-related issues dominated responses regarding barriers (even though perceived lack of software could be viewed as a barrier). For the surveyed companies, the costs of HPC hardware and software are out of reach of many existing budgets, and often the firms do not know enough about HPC to persuade upper management to increase budgets for HPC resources; that is, they are unable to construct convincing ROI arguments. The fundamental cost issue is compounded by the additional cost that would be incurred to hire people who are skilled in exploiting today's difficult-to-use HPC hardware and software.

This study confirmed that an "enabling function" of some kind would be needed to help firms overcome the systemic barriers that are preventing them from using HPC systems. New partnership programs and pricing models may be needed to bring "desktop-only" technical computing users into the HPC fold and to enable more entry-level users to exploit HPC more fully for competitive advantage. About half of the surveyed firms (48%) would be open to paying for consulting help in realizing the technical benefits from HPC – running larger problems than today or running existing problems faster. They would be willing to pay from $2,000 a month to more than $10,000 per month, and they would prefer to pay on a monthly basis. As shown in Figure 4, the most favored outside consulting organizations were "non-profit organizations", "engineering services companies", or "major universities".

2.3.3 ACES Case Study Results

"ACES: Not Ready To Play the HPC Card"

Advanced Computational & Engineering Services, LLC (ACES) is a five-person firm based in Columbus, Ohio that provides a broad spectrum of design engineering services to local and national customers. ACES typically serves as an extension to clients' staff, sometimes providing extra capacity work and at other times contributing new capabilities. The firm currently uses desktop PCs (Windows XP Pro), plus dual-processor Linux and dual-processor Windows servers. The average computational time for a single run of a model is four to eight clock hours, and multiple runs are almost always needed — both to develop the computer model and to run the problem using the model. For one automotive company, ACES needed to run a four-hour problem 50 times in succession. They took advantage of a holiday to schedule the required block of 200 hours. But ACES also has turned away business, such as a recent welding simulation problem, that the company could have accepted if it had HPC capability.

It is not technological unfamiliarity or fear that is preventing ACES from advancing from desktop technical computers to HPC servers—although that is certainly the case with some of "desktop-only" companies surveyed in the studies above. From ACES' perspective, stratospheric software costs are the main barrier to HPC adoption, with hardware costs and concerns about technology obsolescence forming secondary barriers. ***Other than payroll, software is ACES' largest expense today.*** In fact, the company has to ration its use of application software because of the expense. ACES is able to budget for only a limited number of software packages, and for some of these applications the company buys fewer seats than it would like. This rationing sometimes leads to contention for use of certain applications. According to the company, more flexible ISV software licensing schemes would be attractive, such as allowing customers to rent software on a pay-as-you-go, utility basis.

2.3.4 Woodward Case Study Results

"Woodward FST: Software Costs and Finding Experts Are Stalling HPC Adoption"

Woodward FST, a 230-person division of the Woodward Governor Company, builds integrated fuel injection components for commercial and military applications. The company's main products are a range of nozzles that inject and control fuel flow, in order to boost the performance and fuel-efficiency of gas turbine engines used in commercial and military jet aircraft, and land/marine vehicles. Woodward is a major subsystem supplier of Pratt & Whitney and has been selected for Pilot #1 in this effort.

Woodward FST design engineers today use Windows PCs. The software Woodward uses for the computer-aided design (CAD) work includes the Windows XP Pro and Windows 64 operating systems and Unigraph CAD application software. For the subsequent analysis work, the PCs run Windows and Red Hat Linux. The analysis software packages used by Woodward include ANSYS, COSMOS, Star-CD, MATLAB, and SolidWorks.

Woodward FST cannot advance today from desktop computers to more powerful HPC servers because application software for the servers is too expensive, and because the company lacks HPC expertise. Annual licensing fees for the five application software packages the company uses would nearly double if the company began using even a small HPC server. In addition, the firm would need to add staff to the single IT person it employs today as no one internally knows how to use HPC systems. And, while the company expects some progress with the advent of quad-core processors on desktop systems, it notes that quad-core desktops are no less expensive than more capable HPC clusters, and both options still require the addtion of more experienced staff. ***As a result of the being stuck on the desktop, Woodward's business remains successful, but it cannot leverage modeling to improve products.***

Woodward FST is a primary fuel system component supplier to high-performance military and commercial engine aircraft OEMs. Woodward's fuel-system products include the fuel injector for gas-turbine main-burner and augmenter for entire high-performance military engine family such as F100-220, F100-229, F119, F110-129, and F135. The fuel injection system is an essential part of the hot section of the gas turbine engine. The injector provides well-regulated atomized fuel to the combustor and augmenter to sustain the combustion and therefore the thrust of the engine. It is essential that the fuel injection system performs reliably to sustain the normal operation of the engine. The harsh environment in which it operates makes the components' durability and reliability a significant engineering challenge. Failure of the fuel nozzle is a serious mishap in both military and commercial aviation.

Woodward Design Challenges

Supplementary to the case study above, Woodward provided a competition-sensitive white paper of the design challenges it faces routinely as a supplier to P&W. This document summarized below can be made available to the

Government upon request. Fuel injection components constitute three to seven percent of the total engine cost. However, it constitutes up to 1.5% of the engine weight. The fuel nozzle represents a relatively high cost item within the engine system. This is primarily due to the extensive design cycle and complex manufacturing processes. The fuel injector must deliver the fuel to the combustor but it must also protect the fuel before it is delivered. This results in design complexity. Design requirements include precision metering of the fuel, spray performance over wide operational range, thermal protection of the fuel, and structural integrity of the fuel system. The current process does not meet Woodward's needs. Design issues include:

Design Cycle Time. The fuel injector design is potentially a critical path for the system development cycle due to the low level of automation and optimization of simulation and analysis. Lack of accuracy and fidelity in modeling due to limited computing resources also contributes to the bottleneck. These limitations result in excessive design iterations.

Price and Performance. Fuel injector suppliers face key business challenges such as annual price reductions, shorter development cycles, and ever-tightening environmental requirements. Significant engineering efforts must be invested to achieve higher performance fuel injector design. Fuel injectors must now achieve cleaner emissions, weigh less, obtain longer life, have higher reliability and accommodate harsher environments.

Turn-Backs. Turn-backs are tolerated in the course of business, but add cost and waste to the product. Products are not fully optimized for performance or manufacturability. In part, this is a result of inadequate computing resources. During the initial testing and manufacturing, flaws are discovered that result in re-design, rework, and scrap, incurring a great deal of cost. These turn-backs result in schedule and product shipment delays.

Design-Escapes. Design escapes, sometimes found many years into production, result in costly product recalls and field service interruptions, and possibly the loss of the aircraft or the loss of life. These result in intangible, but possibly catastrophic, future costs to the company.

The impact of turn-backs and design-escapes is of particular concern to Woodward and its customers. Woodward provided several competition-sensitive examples of turn-backs and design-escapes not included in the case study report as shown in (Figure 5). Additional documentation on the turn-back examples is available to the government upon request. These examples demonstrate that higher fidelity analysis has been instrumental in root cause investigation of turn-backs and design-escapes, but could be used to make better initial designs. Two impediments to HPC in the design process are: 1) the availability of HPC at the time of design, and 2) the cost of infrastructure and maintenance for HPC facilities. This includes hardware and software costs and staffing of highly skilled engineers. The higher initial and recurring costs of HPC require that the benefit be demonstrated to the decision makers. Lack of availability of HPC capabilities on a "try-and-buy" or "on-demand" basis inhibits the technical community from demonstrating these benefits to management.

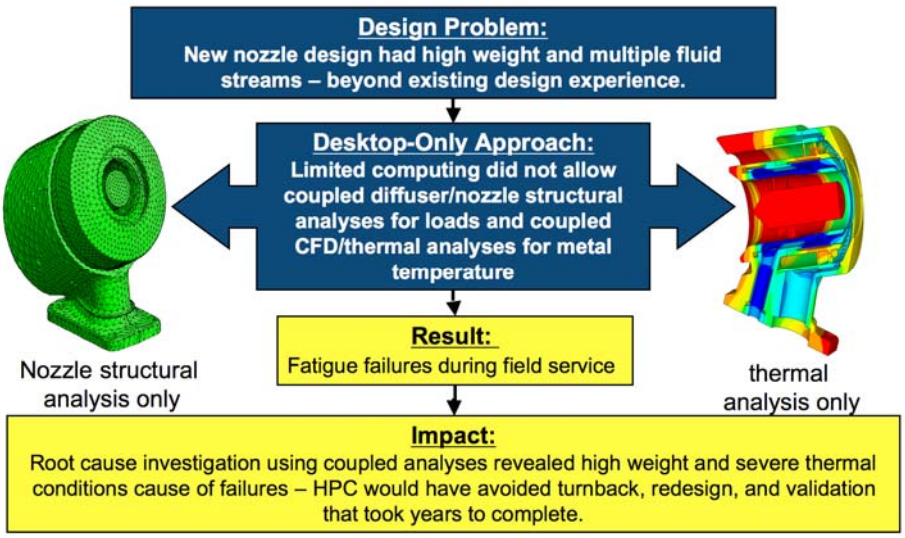
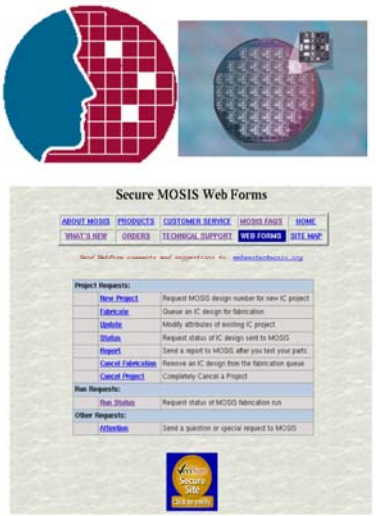


Figure 5. Woodward Example Design Escape Due to Lack of HPC Access

2.3.5 Business and Technical Concept Study Results

"Innovation – Solution Portal to Information, Resources, and Experts (I-SPIRE^{SM})"

One of the early motivations for developing the business concept in *Phase 1* was MOSIS, the Metal Oxide Semiconductor Integration Service. Looking back 25 years, the semiconductor industry was in a situation similar to where HPC is today. Only a select few well-funded companies and researchers could afford to develop computer chips. Given the cost of VLSI fabrication, the chip foundries focused their support on large-volume customers. Lack of talent was a major issue because few teams in the country had the capacity to develop chips. In response to this, DARPA initiated MOSIS, which is a brokerage service that contracted directly with chip foundries for wholesale access to commercial VLSI fabrication in quantity, and then retailed subdivided wafer space on a non-profit basis (Figure 6). The result is an extremely cost effective mechanism for producing prototype and limited production quantities of unique chip designs. The impact has been immense. From 1990 to 2003, over 65,000 students learned chip design through MOSIS. Many commercial chips started as MOSIS prototypes. Today, MOSIS is a self-sustaining non-profit service operating within USC ISI. The DoD still benefits directly from its initial investment in reduced VLSI development cost. The success of MOSIS raised the question: *Why not build a similar brokerage service for HPC?* This initiated the concept for an I-SPIRE^{SM} brokerage service that will contract directly with HPC providers, independent software vendors, and domain experts wholesale and then provide industry-tailored solutions on a "pay-as-you-go" basis.



- **MOSIS has provided 26 years of fabrication brokering services to the nation's VLSI designers**
- **MOSIS is the only non-sponsored service in the world which provides access to multiple technologies & fabricators**
  - Early seed funding provided by DARPA
  - Early VLSI fabrication funding by NSF & DARPA
- **Non-profit organization managed by USC ISI**
- **Academic Outreach (1990 to 2003)**
  - 66,539 students have learned chip design
  - 13,734 designs have been completed
- **Nation's Innovation Engine**
  - Sun Microsystems SPARC, SGI's MIPS, Mobius Microsystems, Many others …

Figure 6. MOSIS, Example of a Government, Industry, Academic Partnership

Although MOSIS provided an inspiration for the I-SPIRE^{SM} brokerage service, the potential HPC market is considerably more complex because of the nature of the end products that are delivered. MOSIS takes a validated VLSI design database as input and delivers packaged computer chips of that design. I-SPIRE is combining software and computing to provide *software as a service* (SaaS). The mechanics of purchasing utility computing HPC cycles and software is similar, but delivery of the services is very different.

The surveys above indicate that the ***primary technical driver*** for desktop-only companies is the availability of "strategic fit" software (both models and applications) that closely matches the specific needs of that company. The potential customers for I-SPIRE do not necessarily have experience with or interest in HPC as a technology. Therefore, HPC must be encapsulated for most customers as a highly tailored software service relevant in the customer's domain. This suggests that, in addition to the brokerage service functions of I-SPIRE, careful attention needs to be paid to the distribution mechanism for the derived services Because of the relative importance of outside organizations to I-SPIRE, the technical and business architecture is broken down into two major components: the *I-SPIRE^{SM} Brokerage Service* and *I-SPIRE^{SM} Access Partners*. The architecture as shown in Figure 7 represents the ***ultimate long-term full-scale vision for the service, of which, a subset will be demonstrated*** in Phase 2. The I-SPIRE^{SM} Brokerage Service on the bottom is the national non-profit entity that negotiates with core services suppliers and provides a marketplace for value-added services built on top of them. The distribution channels, or access partners, on the top half of the diagram are organizations that can reach potential customers using industrial support infrastructure *that already exists at a local, regional, or national scale*.

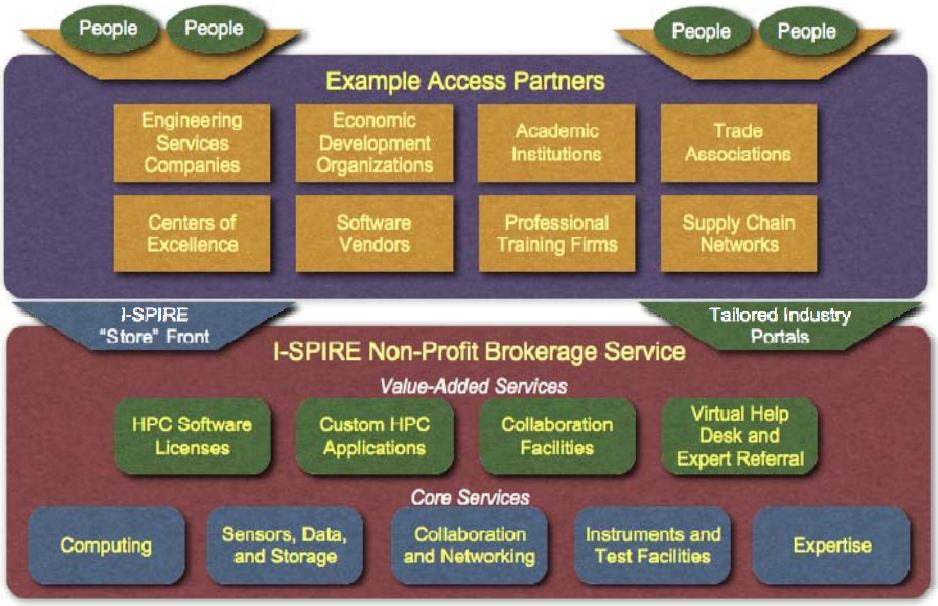


Figure 7. I-SPIRE[SM] Non-Profit Brokerage Service and I-SPIRE[SM] Access Partners

I-SPIRE[SM] Brokerage Service

The I-SPIRE[SM] brokerage service will contract directly with high performance computing providers, independent software vendors, and domain experts wholesale and then retail industry-tailored solutions to companies on a "pay-as-you-go" basis. A good analogy for I-SPIRE is a web hosting service, but focused on technical computing rather than the database functions. The key components of I-SPIRE are:

- ***Core Services*** are the commodity products and services that the brokerage service is buying in quantity from commercial suppliers and reselling as subdivided access directly as a product or indirectly through a value-added service. Core services scale with demand because they build on larger markets. For example, financial services companies are heavy users of utility computing. I-SPIRE[SM] will benefit from price competition and spare capacity generated by this market.
- ***Value-Added Services*** are unique solutions provided to specific classes of customers through independent software vendor back-ends or thin-client web portals. Scaling the value added services is the primary challenge of I-SPIRE, because each industry (and possibly each company) will need software customization requiring both HPC and domain expertise.
- ***I-SPIRE[SM] Store Front*** is a web-based electronic commerce interface for the goods and services available through the brokerage service. This infrastructure will also support the electronic commerce backend functionality for third parties that want to sell custom computing services.
- ***Tailored Industry Portals*** are web portals for highly specific industry solutions as thin-client applications. They require the development of custom HPC applications, domain-specific models, or design templates rather than a general-purpose simulation tool.

I-SPIRE[SM] Access Partners

The most important capability that I-SPIRE will develop is its ability to scale tailored solutions for very specific industry segments. Access partners are organizations that can reach potential I-SPIRE customers using industrial support infrastructure *that already exists at a local, regional, or national scale*. In the web hosting analogy, the access partners are the web designers or database developers that generate a majority of the hosted content. With I-SPIRE, the content consists of tailored industry web portals or stand-alone software front-ends that integrate brokered computation services into customers' design workflows. In some cases, I-SPIRE[SM] will be completely transparent to the customer that is using the service through a third party such as Edison Welding Institute's E-Weld Predictor portal. In other cases, I-SPIRE[SM] will be visible to the end customer that will purchase solutions from the access partners directly through the I-SPIRE[SM] Store Front.

Independent Software Vendors can build I-SPIRE service back-ends directly into their desktop software packages in order to provide a computation acceleration service to their customers. Several technical computing suppliers provide HPC backend interfaces to their desktop products for large customers. I-SPIRE can use this capability.

- ***Engineering Service Companies***. Many firms turn to engineering service companies for access to tools and expertise they lack in-house. One example is ACES, the first case study discussed in Section 2.3.3. If it had metered access to HPC cycles and software, ACES has the expertise to build customized solutions for its customers and pursue business it currently must turn down. Another example of is Edison Welding Institute (EWI). ***The in-depth survey in Section 2.3.2 was of EWI members.*** EWI has the desire to provide turnkey solutions for its customers for more routine welding and joining problems

- ***Supply Chain Networks***. The *principal business motivator* is customer a requirement. This suggests that major system integrators can have a influential role in the adoption of HPC. Woodward is a supplier of engine fuel systems in Pratt & Whitney's supply chain.

- ***Economic Development Organizations.*** State, regional, and national government organizations provide advanced tools and training (or links to them) for companies. Partnerships with these organizations may involve the development of small business innovation research (SBIR) programs that enable the development of industry segment solutions on I-SPIRE**SM** brokered services.

- ***Academic Institutions*** such as community colleges, colleges, and universities provide companies with access to unique tools and expertise through education programs and technology incubators/outreach. An example is California State University Los Angeles (CSULA), which sponsors design clinics for local companies. It provides low-cost design expertise to the companies and mentored hands-on engineering design experience for undergraduate and graduate students. One of the challenges CSULA presently has is that once the design clinic is complete, the HPC machine and software is no longer available to the company. I-SPIRE addresses this problem.

- ***Trade Associations*** exist across most professions and industries and provide linkages to value-added products and services for their members. I-SPIRE will use the Council ties to most of the major industrial trade associations for outreach and marketing of I-SPIRE services. The trade associations can help promote to their membership the value of modeling and simulation with HPC and alert them to the availability of systems and expertise through I-SPIRE.

- ***Centers of Excellence*** are academic or government regional centers with unique instruments, computing facilities, etc. that can benefit from or provide additional I-SPIRE services. For example, while interviewing the engineering firm ACES for the case study in Section 2.3.3, they recommend that they would also like reliable remote access to an electron microscope service where the materials stress analysis sample from physical prototyping could be shipped to an instrument technician and then could be examined in real-time over the Internet. I-SPIRE could broker access to these non-computational facilities and provide scheduled access. I-SPIRE will use the Council's deep relationships with its broad university membership and the national laboratories for outreach and marketing of I-SPIRE services.

- ***Professional Training Firms*** can use I-SPIRE for relevant training programs on virtual prototyping and large-scale data analysis. Often, software vendors provide firms with free licenses for training purposes. In some cases, the professional training firms *are* the software vendors.

*I-SPIRE**SM** Technical Architecture*

The I-SPIRE**SM** broker shown in Figure 8 is essentially a web server that provides remote access to computing and software resources through an electronic commerce storefront, industry-tailored thin-client web application interfaces, or web services remote procedure calls from commercial software packages. I-SPIRE will utilize service-oriented architecture (SOA) standards and tools popularized in what is known as the "Web 2.0" revolution, which enables highly interactive web applications. On the back-end of this web-based system are custom interfaces to utility computing vendors, software vendors, and solution providers of various "value-added services".

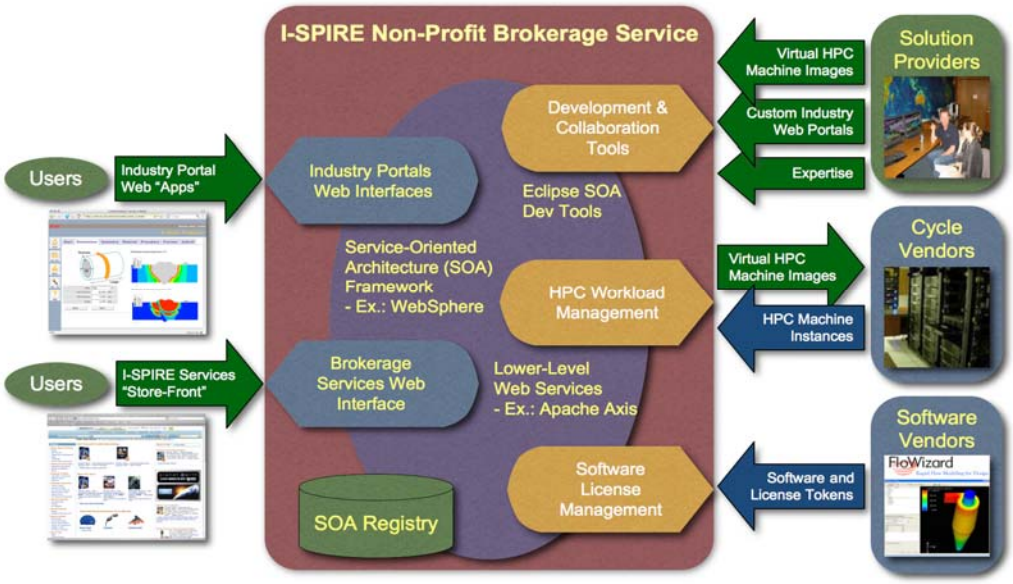


Figure 8. I-SPIRE[SM] Technical Architecture

The technical architecture of the brokerage service has several major components:

For lower-level web services, the brokerage server will use Apache Axis, an open-source implementation of SOAP ("Simple Object Access Protocol"), which is an XML-based protocol for exchanging structured information. Axis has the advantage of being highly scalable, since it is used successfully in very large web server deployments.

The SOA framework will utilize a commercial or open-source package such as IBM's WebSphere®, which has standard plug-in modules for electronic commerce storefronts, metering and billing, application session management, registry data management, and security and authentication. The tools for web application development and supporting business services are very robust. I-SPIRE will take advantage of existing frameworks to develop a federated service registry to manage all of the supplier and customer information related to the brokerage service.

The interfaces to solution providers will be built on the Eclipse IDE framework. This open-source integrated development environment has extensive capabilities for building and interfacing to web services applications. Eclipse supports third-party plug-ins. I-SPIRE will develop specific plug-ins that enables solution providers to use the brokerage services' resources from their stand-alone applications. Many independent software vendors use Eclipse in their product development.

I-SPIRE will utilize virtual machine technology for configuration management. Software applications will be distributed and managed as virtual machine images. Virtual machine instances of these applications will be booted "on-demand" within the HPC cycle vendor's facilities.

2.4 Conclusions

The HPC-ISP team reviewed the I-SPIRE[SM] business concepts with the McDonough School of Business at Georgetown University. The concept was presented to two business professors in March 2007:

Dr. Betsy Page Sigman, teaches Management Information Systems, Electronic Commerce, and Database Development and Management. She has expertise in SAP business software. Her experience includes Senior Project Director for a major market research firm, Wirthlin Worldwide.

Dr. Jeffrey T. Macher, is an Assistant Professor at Georgetown. He teaches Strategic Management, Technology and Knowledge Management, and Microeconomics. His expertise is in ecommerce, vertical specialization in the semiconductor market, and globalization of innovation.

The overall feedback was that I-SPIRE benefited from good timing from an information technology perspective (e.g. Web 2.0 services and SOA). The I-SPIRE approach is consistent with successful web businesses in other domains. The "mash-up" concept of pre-negotiated agreements with service providers is a successful model.

Specific feedback included:

What is the best business model for value-added services? The first recommendation was a "business-to-business" or "web hosting" model, where I-SPIRE provides a fixed set of hosted core services, charges the "access partners" directly for what they use, and then lets the "access partners" charge customers directly. The second recommendation was the "Google Adsense" model, where I-SPIRE provides tools so that "access partners" can host tailored solutions on the brokerage Service. I-SPIRE charges end-customers directly for resources consumed. I-SPIRE would then pay the access partners that build value-added services a commission based on the resource-hour consumption they generate. These well-recognized business models have the advantage that the "access partners" have a financial incentive to participate in I-SPIRE.

What actions are required to avoid lack of adoption? The team should focus on getting core access partners and service providers on board early to accelerate customer outreach and minimize customer support expense. It is necessary to create a market-driven industry tailored portal development community to provide a broad spectrum of readily available industry specific application portals. The core project should provide tools and not the "content" (applications). The team needs to focus on a few targeted industry segments in the beginning. Manufacturing segment alone is too broad. ***It is important to implement early focused regional pilots to provide early "success stories" funded by federal or state agencies. These pilots can then form the basis for the launch of a sustainable non-profit brokerage service.***

This project produced a significant body of evidence that has been used to motivate the adoption of high performance computing throughout the U.S. manufacturing sector. The reports generated under this effort have been used to guide national policy driving U.S. innovation.

2.5 Recommendations

Guided by the surveys, case studies, and business studies of the first phase, the research team recommended the development of an I-SPIRE℠ brokerage service prototype and three market segment pilots prior to the launch of a national non-profit service. The goal of this effort will be to establish relationships with suppliers and access partners, develop and validate functional interfaces, and generate critical "success-stories" for the future I-SPIRE launch. The proposed *Phase 2* was organized into four tasks:

Task 1 – I-SPIRE Strategic Direction will establish strategic relationships with brokerage service suppliers and key access partners, and also develop a sustainable business plan for I-SPIRE in preparation for the future launch of the non-profit brokerage service.

Task 2 – I-SPIRE Brokerage Service Engineering will construct the prototype brokerage service, which will develop and validate functional interfaces to the computing and software vendors and tools for collaboration and development with "value-added" services.

Task 3 – Market Pilots will develop the three industry-focused pilots that explore different interface mechanisms, such as thin-client web interfaces, thick-client simulation tool interfaces, and data analysis language interfaces, as well as different access channels to potential customers, such as engineering services companies, independent software vendors, and DoD supply chain networks. The market pilots include:

- ***Pilot #1: Software Services Market Pilot*** includes interactive access for interpreted analysis languages (Mathematica, MATLAB, or Python) using Interactive Supercomputing's Star-P and Wolfram Research's gridMathematica® tools. The key metric is response time for small, short duration (seconds to minutes), interactive multi-user jobs.
- ***Pilot #2: Engineering Services Market Pilot*** includes Edison Welding Institute (EWI), a leading expert in materials joining and welding. EWI and Ohio Supercomputer Center (OSC) will integrate their prototype E-Weld Predictor™ service with I-SPIRE℠. The key metric for this pilot is resource utilization for medium duration (minutes to hours), web-dispatched jobs using a travel-like reservation system for HPC resources.

- ***Pilot #3: DoD Supply Chain Market Pilot*** includes Pratt & Whitney and Woodward, a "desktop-only" fuel systems supplier in their supply chain. This pilot will accelerate desktop simulation using brokered HPC cycles and licenses for extended-capability multi-physics processing jobs. The key metric is application speedup for complex, long duration (hours to days) batch jobs.

Task 4 – DoD Supply Chain Competitiveness and National Outreach will examine competitiveness issues within the DoD supply chain, publish success stories of DoD suppliers that are realizing benefits from using HPC, and undertake a broad-based, multifaceted outreach program to "get the word out" about the critical importance of modeling and simulation with HPC to defense industrial strength.

These recommendations were presented to Dr. Tether, DARPA Director, and were subsequently modified with his feedback. The next phase of this study was continued by DARPA under the High Performance Computing Innovation Service Portal Pilots (HPC-ISP-PILOTS).

3. Computer Architecture Industry-Academic Consortium (CAIAC) Study

HPC-ISP-CAIAC was a six-month study led by ISI to investigate long-range research into parallel systems, software, and applications for next generation computing devices.

3.1 Introduction

The microprocessor hit an historic inflection point in 2005 that has changed the microprocessor industry. The industry is no longer pursuing evolutionary changes to single-core architectures in tandem with increasing clock speed as its primary technical or economic growth path. Power efficiency limitations and the difficulty of extracting additional parallelism from a sequential instruction stream now prevent the industry from maintaining simple binary compatibility while increasing clock speeds. The industry has been forced to adopt a parallel computing model to continue to benefit from the potential performance benefits brought about by increasing transistor densities enabled by Moore's Law.

While the exploitation of parallelism does provide the potential to continue performance improvement as Moore's Law continues, this move to parallelism provides challenges for both hardware and software design. Hardware questions include the composition of the parallel cores, which may be homogeneous or heterogeneous, interconnect, and memory organization. Software questions center around how to make these parallel architectures programmable. Parallel programming is not new, but it has always been practiced in niche communities, such as high performance computing, that have highly trained, specialized "hero" programmers. If multi-core architectures are going to propel continued growth of the microprocessor industry, programming paradigms that make parallel programming "available to the masses" will be needed. If the problem of parallel programming for multi-core is not solved, it could mean disaster for the microprocessor industry, and, in turn, for the DoD, which has grown dependent on exponential improvements in processing for many applications. This potential disaster for the microprocessor industry is also an opportunity for computer architecture, which has been constrained by legacy sequential processing models for decades.

3.2 Methods, Assumptions, and Procedures

In this study, ISI solicited and summarized the technical recommendations from three regional academic workshops of leading professors from computing-related fields assembled to investigate this chip-level parallelism challenge. Dave Patterson from University of California, Berkeley led the Silicon Valley Academic Workshop, which was held in August 2007 at Stanford University. Doug Burger from the University of Texas at Austin led the Austin Academic Workshop in September 2007. Finally, Vivek Sarkar from Rice University led the Boston Academic Workshop held at MIT in October 2007.

In addition to the academic workshops, Robert Graybill from ISI brought together DARPA representatives with leading industry executives from computing vendors, software vendors, and users at a meeting in San Mateo, CA in December 2007. He presented a potential industry-government consortium approach similar to the very successful Focus Center Research Program (FCRP) managed by the Microelectronics Advanced Research Corporation, a subsidiary of the Semiconductor Research Corporation (SRC). Like the FCRP program, the Parallel Systems, Software, and Applications Research (PSSAR) program could be a membership organization administrated by SRC to address pre-competitive high-risk innovative research by utilizing the U.S. research university system. Industry members and the government would match contributions and share in the research products and resulting intellectual property in a similar fashion to FCRP participants. PSSAR is designed to give the U.S. preferential access to technology emerging from the program. DoD scientists and engineers as well as defense contractors would be involved in all stages of this research.

3.3 Results and Discussions

The consensus of the participants was that the computer industry is rapidly approaching a crisis where the future growth of processor performance for many applications is uncertain. The participants warned that a major research investment is required to "break the habit of von Neumann sequential computing models" in order to derive performance growth from chip-level parallelism. They recommended that the government and industry fund "diverse academic research efforts" that are able to break traditional hardware and software boundaries in the development of new parallel software systems. Using this guidance, ISI brought together DARPA representatives with leading industry executives from computing vendors, software vendors, and computing users to explore the potential of industry-government consortium approaches for funded focused research in this area. A concept paper, entitled *Parallel Systems, Software, and Applications Research (PSSAR) Consortium Concept Paper*, was submitted.

3.4 Conclusions

The development of software and hardware of multicore systems of the future are interdependent. The hardware must be designed to support the new system software stacks and applications, and the software must be designed to take advantage of the hardware. It will be critical for multi-disciplinary teams that include hardware, software, and applications experts to carry out this research. These teams also must be large enough to develop systems that are credible alternatives to existing processor-based systems. It will be impossible to credibly evaluate new architectural approaches without developing complex, large scale systems that include software and hardware.

It is also critical that significant investments are made in multiple approaches. There are no obvious or evolutionary approaches to solving the challenges of multicore software and hardware. While investments in focused, unified approaches are needed to develop system stacks to allow full evaluation, it is too risky to focus all research investments in one or two approaches. It is necessary to fund a large program that funds multiple interdisciplinary teams, each team focused on a single, coherent approach. While we believe the creativity required to solve these challenges means that each team should have freedom to pursue their own approach, there is still experimentation and evaluation infrastructure that can and should be shared between teams.

3.5 Recommendations

We believe that there should be a diversity of research efforts examining different solutions supplemented by individual point studies that together examine different solutions to this grand challenge. Each effort will consist of a team of researchers and experts across many layers of the traditional system stack, including algorithms, applications, programming languages, compilers, runtimes, libraries, tools, operating systems, and architecture. Each effort must be large enough to allow serious, integrated, at-scale research to be completed and long enough in duration to make substantial progress on this difficult challenge. These efforts must have long-term funding continuity to enable ambitious new approaches to be explored. Five years is a minimum duration; longer would be better. The intent is to allow researchers to develop enough novel technology to reach escape velocity from the gravity well of existing conventional computing system architectures. Given the need to rework many of these system interfaces will involve interactions with experts in different domains; it will be important for many of the researchers from each team to be physically collocated, maximizing the interaction among them.

To permit comparability across the various efforts, the teams should focus on a common set of problems (although each team may focus on some additional problems that are unique to their project). The applications must be of general enough interest to ensure that if substantial progress is made on them, the research output will be eagerly adopted by the mainstream microprocessor industry. The major application classes initially under consideration are natural language processing, data mining, image recognition and search, and simulation and modeling of interactive virtual worlds.

The teams should develop at-scale prototype systems that demonstrate dramatic gains across the target applications. These prototype systems will be full systems, with software written at many layers. We expect that systems may be emulated using FPGAs, existing multi-core systems and Time Transition Model (TTM) based modeling solutions with perhaps limited development of hardware boards and/or simple ASIC test chips. It is critical that these efforts support large-scale prototype software systems building. The prototype systems could also provide a foundation for reference implementations for industry standards that may emerge from this program. The teams will reuse and share infrastructure to the extent that it accelerates their research. If enough commonality across the teams exists, it may be fruitful to provide support for maintenance of cross-project infrastructure.

It is well known that efforts to develop parallel computing for the general-purpose computing community have yet to produce viable tools for wide-spread use. The lack of tools is of particular concern today since the era of the “killer micro” is over. Parallel architectures in the past were competing with sequential microprocessors that doubled in performance every 18 months at a fixed cost. This exponential increase in performance drove the IT industry and made it nearly impossible for other technologies to compete. Now, sequential microprocessor performance has essentially stopped increasing, and the community must and are finding some alternative approaches with new challenges. An example is the emergence of the “killer GPUs” and the need to support efficient heterogeneous computing environments.

We have some reason for optimism, since new technologies have emerged that change the possibilities of exploiting parallelism. On-chip bandwidths and latencies are orders of magnitude better than the inter-chip bandwidth and latency of many-socket systems, and new software technology such as just-in-time (JIT) compilation makes it feasible that software can adapt to growing parallelism. In addition, there are improvements in research

infrastructure that enables at-scale research that was not feasible in the past. The open source movement allows thousands of developers to contribute to code development, and dense field-programmable gate array (FPGA) technology allows the emulation of many core systems, providing orders of magnitude performance improvement over software simulation. The confluence of necessity and enabling technologies makes now the time to develop novel single-socket (multi-core) computing systems that will replace the sequential microprocessors that have provided exponentially increasing performance over the last several decades.

In summary, it is an enormous challenge to build systems that can exploit scalable, energy-efficient parallelism and yet can also be programmed effectively by a broad community of programmers. The benefits of meeting this challenge, of course, are commensurately high: at least two orders of magnitude improvement in power/performance, another 15 years or more of fruitful silicon scaling, continuity of economies of scale that permit successive CMOS manufacturing investments, and new, exciting classes of applications becoming feasible, providing benefits to science, health, business, and society. To get there will require attracting the best minds in the field and to enabling them and their teams to work hard and creatively on this longstanding problem.

Potential Organizational Structure

The Semiconductor Research Corporation (SRC) was founded in 1982 at the recommendation of the board of the Semiconductor Industry Association (SIA) to ensure continued research on advancing silicon technology and improved manufacturability of integrated circuits (ICs). SRC operates multiple research programs both in the U.S. and globally to provide competitive advantage to its members as the world's premier university research management consortium. The SRC provides low-overhead research and related programs that meet the needs of the semiconductor industry for technology and talent. SRC maintains core competencies in contracting and management for research at universities with terms and conditions that assure members' ability to use results with minimal risk of future encumbrances, including intellectual property protection of significant research results and mechanisms for timely transfer of research products. In addition, SRC manages student programs to maximize the flow of talent to the member companies. SRC obtains funding for its programs by recruitment and retention of member companies through appropriate fee algorithms for the relevant market segments and by leveraging government funding.

In 1997, SRC formed a subsidiary, the Microelectronics Advanced Research Corporation (MARCO), to support the Focus Center Research Program (FCRP), an initiative for pre-competitive, cooperative, long-range, applied microelectronics research sponsored by members of the U.S. semiconductor industry, U.S. supplier industry, U.S. EDA industry, and the U.S. Department of Defense (DoD). The FCRP concentrates resources on those areas of microelectronics research that must be addressed to maintain the historic productivity growth curve of the industry, strengthen the university research infrastructure and expand its capabilities in silicon related research, achieve critical mass through relatively large blocks of funding together with the active participation of industrial visiting scientists, and provide the optimal balance of creative freedom and targeted objectives. The FCRP has been an extremely successful program for industry, government, and academia. The SRC received the 2005 National Medal of Technology®, which is the nation's highest honor for technical innovation. SRC is recognized for constructing the world's largest and most successful university research force in support of the rapidly growing semiconductor industry; for proving the value of collaborative research as the first high-tech research consortium; and for creating the concept and methodology that evolved into the International Technology Roadmap for Semiconductors.

It may be feasible to replicate the successful FCRP model for a new program called Parallel Systems, Software, and Applications Research (PSSAR). A new subsidiary of SRC, PSSAR Corporation (PSSARC) might be established to create a unique legal entity for a new governing board representing industry members including computing vendors, software vendors, large computing users and government partners including DARPA In addition, PSSARC could provide a legal vehicle for shared intellectual property through consortium membership agreements. PSSAR would have a similar structure and organization to the very successful FCRP program. However, its specific features would reflect its industry and government member requirements and the distinct research objectives of the new program.

The PSSAR program could be funded by industry membership fees and matched funding from DARPA. The industry fee structure would require additional negotiation with the founding members. Fees in the FCRP program are currently computed as a percentage of relevant revenue. The PSSAR program would consider a similar formula, possibly a percentage of relevant R&D. SRC would administer PSSAR and an appointed technical director would monitor the research teams, guide inter-team collaboration, and recommend infrastructure consolidation where appropriate. A governing board of industry and government members would oversee PSSAR.

The PSSAR program would address technology research by utilizing the U.S. research university system to work on long-range innovative research. Research would be solicited through a joint research announcement between DARPA and PSSARC. Successful proposals would be expected to identify potential core set of potential breakthroughs and team that provides reason to believe that a project can be successful. Only U.S. universities would be considered for funding. Proposals may include other organizations, but those other organizations must be funded by some other means, they cannot receive funding for activities through the proposed effort. Research contemplated is generally high-risk, high-payback, long-term and targeted technology with a 5-10 year horizon. Each funded research project would require compelling and well-articulated vision for long-term, exploratory research important for U.S. competitiveness; a vision that is both embraced and enthusiastically deployed; a vision that leads to the conception, demonstration, and evaluation of revolutionary systems. Research selection panels would include both government and industry personnel.

4. References

[1] Council on Competitiveness, “Gaining New Ground: Technological Priorities for America’s Future”, http://www.compete.org/hpc, 1991.

[2] Council on Competitiveness, “Endless Frontiers; Limited Resources”, http://www.compete.org/hpc, 1996.

[3] Council on Competitiveness, “Imperatives for Innovation”, http://www.compete.org/hpc, 2001.

[4] Council on Competitiveness, “Innovate America: Thriving in a World of Challenge and Change”, http://www.compete.org/hpc, 2004.

[5] Council on Competitiveness, “Final Report from the First Annual HPC Users Conference: Supercharging U.S. Innovation & Competitiveness”, http://www.compete.org/hpc, 2004.

[6] Council on Competitiveness “Survey of Industrial HPC Users”, http://www.compete.org/hpc, 2004.

[7] Council on Competitiveness, “Final Report from the Second Annual HPC Users Conference: Accelerating Innovation for Prosperity”, http://www.compete.org/hpc, 2005.

[8] Council on Competitiveness, “Council on Competitiveness Survey of ISVs Serving the High Performance Computing Market – Part A: Current Market Dynamics”, http://www.compete.org/hpc, 2007.

[9] Council on Competitiveness Survey of ISVs Serving the High Performance Computing – Part B: End User Perspectives”, http://www.compete.org/hpc, 2005.

[10] Council on Competitiveness, Workshop Report, Accelerating Innovation for Competitive Advantage: The Need for HPC Application Software Solutions”, http://www.compete.org/hpc, 2005.

[11] Council on Competitiveness, “Grand Challenge Case Studies”, http://www.compete.org/hpc, 2005:

a. Keeping the Lifeblood Flowing: Boosting Oil & Gas Recovery From the Earth
b. Full Vehicle Design Optimization for Global Market Dominance
c. Auto Crash Safety: It’s Not Just for Dummies
d. Customized Catalysts to Improve Crude Oil Yields: Getting More Bang from Each Barrel
e. Spin Fiber Faster to Gain a Competitive Edge for U.S. Textile Manufacturing

[12] DIHT Newsletter, Information Directorate, Air Force Research Laboratory, Rome, NY.

[13] Council on Competitiveness, “REVEAL: Council on Competitiveness and USC-ISI Study of Desktop Technical Computing End Users and HPC”, http://www.compete.org/hpc, 2007.

[14] Council on Competitiveness, “REFLECT: Council on Competitiveness and USC- ISI In-Depth Study of Technical Computing End Users”, http://www.compete.org/hpc, 2007.

5. Acronyms

Acronym	Definition
ACES	Advanced Computational and Engineering Services
AFRL	Air Force Research Laboratory
AMD	Advanced Micro Devices, Inc.
ARO	After Receipt of Order
ASIC	Application-Specific Integrated Circuit
CAIAC	Computer Architecture Industry Academic Consortium
CFD	Computational Fluid Dynamics
CMOS	Complementary Metal-Oxide Semiconductor
CoC	Council on Competitiveness
CPU	Central Processing Unit
DARPA	Defense Advanced Research Projects Agency
DoD	Department of Defense
EDA	Electronic Design Automation
EWI	Edison Welding Institute
FCRP	Focus Center Research Program
FEA	Finite Element Analysis
FPGA	Field Programmable Gate Array
FST	Fuel Systems Textron
GPU	Graphics Processor Unit
GU	Georgetown University
HP	Hewlett Packard, Inc.
HPC	High Performance Computing
HPC MOD	High Performance Computer Modernization
HPC-ISP	High Performance Computing – Innovation Service Portal
I-SPIRE	Innovation – Solution Portal to Information, Resources and Experts
IBM	International Business Machines
IC	Integrated Circuit
ISI	Information Sciences Institute
ISV	Independent Software Vendor
IT	Information Technology
JIT	Just-In Time
LLC	Limited Liability Corporation
MARCO	Microelectronics Advanced Research Corporation
MIT	Massachusetts Institute of Technology
MOSIS	Metal Oxide Semiconductor Integration Service
NICE	National Innovation Collaboration Ecosystem
NSF	National Science Foundation
OSC	Ohio Supercomputer Center
P&W	Pratt & Whitney
PC	Personal Computer
PSSAR	Parallel Systems, Software, and Applications Research
R&D	Research & Development
ROI	Return on Investment
SaaS	Software as a Service
SAP	Service Access Point
SGI	Silicon Graphics, Inc.
SIA	Semiconductor Industry Association
SOA	Service-Oriented Architecture
SRC	Semiconductor Research Corporation
TI	Texas Instruments, Inc.

Acronym	Definition
TTM	Time Transition Model
U.S.	United States
UC	University of California
UCB	University of California, Berkeley
UCSD	University of California, San Diego
USC	University of Southern California
UT	University of Texas
VLSI	Very Large Scale Integrated Circuits